

# **Perspective Research | Methods and Initial Results**

**Yuqing (Candice) Zhang**

## **Introduction**

Currently there have been many articles talking about the determining factors of businesses' success. However, there aren't that many research that studies why businesses' fail. Previous studies found that restaurant failure rates are higher in U.S. Postal ZIP codes where there is a high concentration of restaurants. They also highlighted personal reasons, like poor service and food quality for restaurant failure.

Building on previous studies, I will explore how do the mechanisms of food, service and location impact the closing of restaurants, in Las Vegas area. The dataset I will use consist of reviews from the website Yelp.com. Given a set of customer reviews of a restaurant, the task involves four subtasks: (1) Filtering the raw reviews from Yelp so that we only analyze reviews within our matter of interest; (2) identifying features of the product that customers have expressed their opinions on and find out those that talk about location, food and service; (3) for each feature, identifying review sentences that give positive or negative opinions; and (4) producing a summary using the discovered information.

## **Data**

For this research, I will use the data set that contains information on reviews and businesses from Yelp.com to explore the interplay of my three matter of interest on restaurant close.

This year is Yelp's ninth round of its own dataset challenge. Yelp released a set that includes information about local businesses in 11 cities across 4 countries. The data contains 4.1 million

## **Perspective Research | Methods and Initial Results**

### **Yuqing (Candice) Zhang**

reviews and 947,000 tips by 1 million users for 144,000 businesses. They are enclosed in 5 json files provided by Yelp.com. For my own convenience, I converted all the json files to csv format files.

### **Methodology**

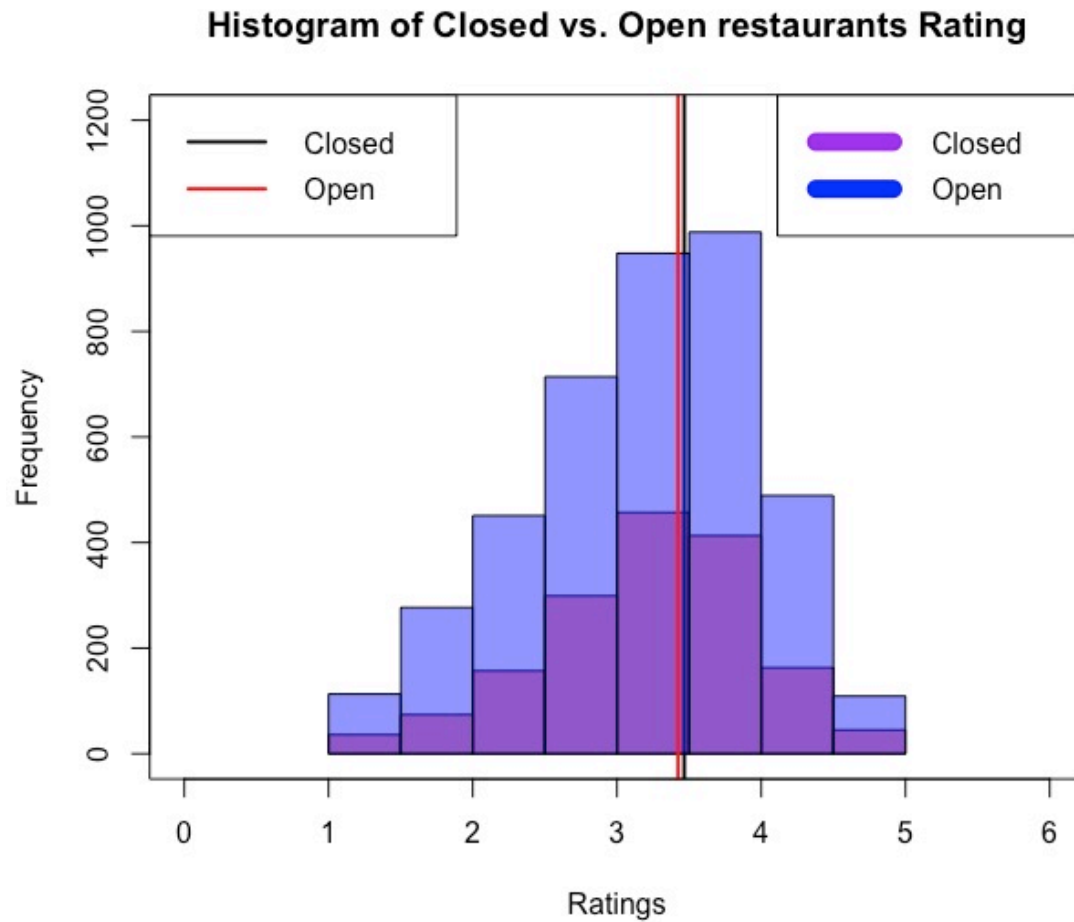
Our task is performed in four main steps:

- 1) Filtering the businesses so that we include only the restaurants in Las Vegas area. I split the restaurants into open and closed restaurants for further comparison. Next sub-step is to split the reviews into year 2012, 2013, 2014, 2015, 2016. I will do the text analysis for reviews on each year and see the pattern.

I made a histogram to compare the ratings for open and closed restaurants. It surprised me that the mean rating of the open restaurants is slightly lower than that of the closed restaurants, indicating the rating might not be a good indicator for predicting whether a restaurant will fail or not.

## Perspective Research | Methods and Initial Results

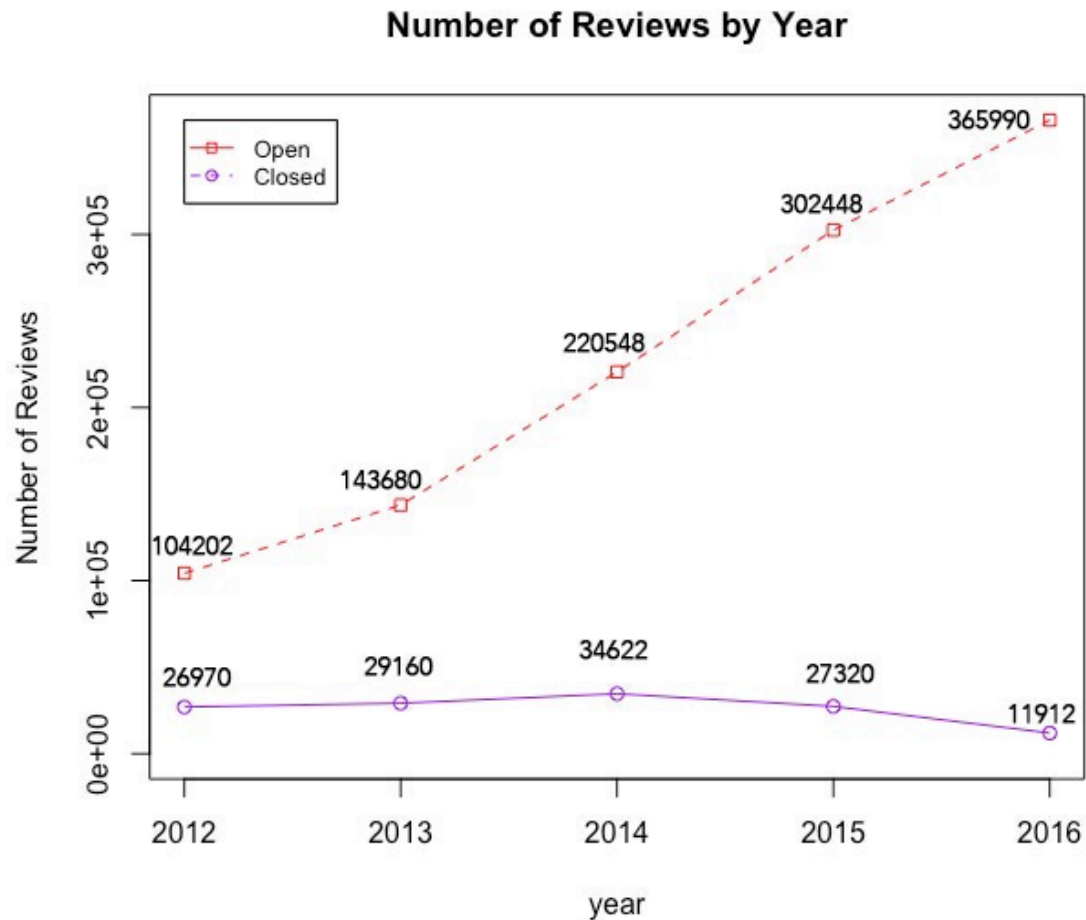
Yuqing (Candice) Zhang



I made another plot to compare the number of reviews by year for open and closed restaurants.

## Perspective Research | Methods and Initial Results

Yuqing (Candice) Zhang



Despite the fact that there are more open restaurants than closed restaurants in Las Vegas: 4089 vs. 1644, the number of reviews for open restaurants is increasing whereas it's relatively steady for that of closed restaurants.

- 2) Mining product features that have been commented on by customers. I will first create three lists of words that pertain to either 'food', 'location' or 'service'. To perform this task I will use NLTK WordNet, as well as Thesaurus to find the synonyms. Then for each restaurant, I will split the reviews into sentences and in the end I will divide each sentence into one of the 3 groups, 'food', 'location' or 'service'.

## Perspective Research | Methods and Initial Results

### Yuqing (Candice) Zhang

3) Next step is Sentiment analysis. This step will be further split to these sub-steps:

3.1. Part-of-Speech Tagging (POS): We will use the NLProcessor linguistic parser to parse each sentence and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc). Each sentence will be saved in the review database along with the POS tag information of each word in the sentence.

3.2. Opinion Words Extraction: Previous work on subjectivity has established a positive statistically significant correlation with the presence of adjectives. Thus the presence of adjectives is useful for predicting whether a sentence is subjective, i.e., expressing an opinion. And therefore this research will use adjectives as opinion words.

3.3. Orientation Identification for Opinion Words: For each opinion word, we need to identify its semantic orientation, which will be used to predict the semantic orientation of each opinion sentence. For this task we will use SentimentIntensityAnalyzer function from NLTK to calculate a positivity score for each opinion word.

3.4. Predicting the Orientations of Opinion Sentences: In general, we use the dominant orientation of the opinion words in the sentence to determine the orientation of the sentence.

4) For each of the three features, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. I will add up the positivity scores for all the sentences in one feature. In the end, I will generate a plot, putting the scores for each year of these three features together.

The above steps will be done separately for open and closed restaurants. The tentative result will be that the scores of all three features for open restaurants will be higher than those of

## Perspective Research | Methods and Initial Results

**Yuqing (Candice) Zhang**

the closed restaurants.

### Reference:

1. C. Bird, (2015, May). Which local business reviews are better: Yelp or Google? [Online]  
Available: <http://localvox.com/blog/local-business-reviews-yelp-or-google/>
2. A. Hicks, S. Comp, J. Horovitz, M. Hovarter, M. Miki, and J. L. Bevan, Why people use Yelp.com: An exploration of uses and gratifications, *Computers in Human Behavior*, vol. 28, no. 6, pp. 2274-2279, 2012.
3. Parsa, H. G., Self, J., Sydnor-Busso, S., & Yoon, H. J. (2011). Why Restaurants Fail? Part II - The Impact of Affiliation, Location, and Size on Restaurant Failures: Results from a Survival Analysis. *Journal of Foodservice Business Research*, 14(4), 360-379. doi: 10.1080/15378020.2011.625824
4. Luca, Michael. "Reviews, Reputation, and Revenue: The Case of Yelp.com." Harvard Business School Working Paper, No. 12-016, September 2011. (Revised March 2016. Revise and resubmit at the *American Economic Journal - Applied Economics*.)